Copyright © 1979 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 9, November 1979
Printed in U.S.A.

The Effects of Several Transmission Systems on an Automatic Speaker Verification System

By C. A. McGONEGAL, A. E. ROSENBERG, and L. R. RABINER

(Manuscript received May 29, 1979)

In an earlier report, the effects of several transmission systems on speaker verification by human listeners were investigated. It was shown that the transmission system played a significant role in the speaker verification process. In this paper, we show the effects of the transmission system on an existing automatic speaker verification system in which the measured features are pitch and gain as a function of time for a specified utterance. In this experiment, there were 10 male and 10 female customers and 40 male and 40 female impostors. Fifty utterances were recorded using a conventional telephone connection over a period of two months. All utterances were post-processed by an ADPCM *coding system and* LPC *vocoding system. When the reference and test utterances were subjected to different transmission systems, no significant difference in the verification accuracy of this automatic system was found. This result verifies that pitch and gain are robust features for use in a speaker verification system.*

I. INTRODUCTION

The automatic speaker verification problem has two aspects—the creation of a reference pattern and the determination of similarity between a test and a reference pattern. When verification is performed over dialed-up telephone lines, the transmission system used in the telephone plant is an additional factor that must be considered. In a recent subjective experiment,[1] the effects of adaptive differential pulse code modulation (ADPCM) coding and linear predictive vocoding (LPC) on the speaker verification accuracy of human listeners was investigated. It was shown that the verification task was easiest (most accurate) when homogeneous systems were used (i.e., the test and

reference utterances were transmitted over the same system) and significantly more difficult (higher error rate) when mixed systems (i.e., different transmission systems) were used for the test and reference utterances. In this paper, we investigate how these same transmission systems affect machine verification accuracy using a system that has been studied for the past few years at Bell Laboratories.[2-6]

The automatic system is based on the analysis of fixed sentence-long utterances in which the verification features are the time variations (contours) of the pitch and gain (intensity) of the utterance. A training set of utterances (both customer and impostor) is required to establish a reference pattern and to choose weights and measurements for the verification process. Following time alignment of the reference and test contours, a combination of weighted Euclidian distances between a set of test and reference measurements is compared with a threshold to determine whether to accept or reject an identity claim.

In an extensive investigation of the automatic speaker verification system over dialed-up telephone lines, Rosenberg obtained an average verification accuracy of about 91 percent.[2] Rosenberg also found that some talkers tended to perform significantly worse than average, and some significantly better than average.

To investigate the behavior of the automatic speaker verification system on different transmission systems, a new data base of customer and impostor utterances was created. Dialed-up telephone connections were used in all recordings. Since the earlier work on human verification used wideband, high-quality recordings, the experiments with human listeners were repeated using the new data base. Following this, a series of experiments was run with the automatic verification system.

The key results of this study are:

(*i*) Human verification accuracy on the telephone speech was essentially the same as previously reported for the high-quality speech, i.e., the highest verification scores were obtained when the reference and test utterances were transmitted over the same system, and significantly lower verification scores were obtained when different transmission systems were used for the test and reference utterances.

(*ii*) Machine verification accuracy on the telephone speech was essentially independent of the transmission system used for the test and reference utterances.

These results tend to confirm the notion that pitch and gain are robust features for verification and hence are suitable for many applications.

The organization of this paper is as follows. The automatic speaker verification system is described in Section II. Section III describes the experimental procedure used to evaluate the automatic verification system. This section includes a description of the speech transmission

systems, as well as the data base used in the evaluation. The machine verification and the human verification results are presented in Sections IV and V. Finally, in Section VI the main results of the experiment are discussed.

II. THE AUTOMATIC SPEAKER VERIFICATION SYSTEM

Although the operation of the automatic speaker verification system has been described previously,[2-6] a brief review is given here. A block diagram of the overall verification system is shown in Fig. 1. Two inputs are provided to the system. These are an identity claim which retrieves reference data associated with the claimed identity and a sentence-long sample utterance. The sample utterance is analyzed to extract time functions or contours of specified features which are compared with (previously obtained) reference contours. Reference contours are obtained by averaging and combining sets of contours obtained from training utterances from the individual whose identity is claimed. The features used in this experiment are the intensity (gain) and pitch period. The gain contour is normalized so that its peak

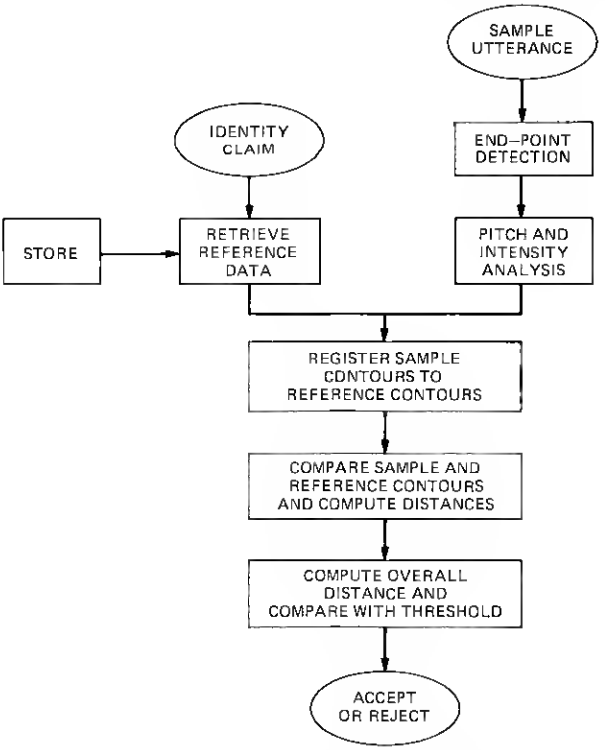

Fig. 1—Flow diagram of the verification system.

over the entire utterance is a fixed value and is low-pass filtered to give a smooth contour.

Before comparing sample and reference contours, time registration is carried out. Using a dynamic programming technique, the sample intensity contour is time-warped so that corresponding events in the sample and reference contours are aligned in time. The resulting time-warping function is also applied to the sample pitch period contour in order to align it to the reference contour. Following registration, the contours are divided into 20 equal length segments. In each segment, a set of measurements is applied to both the sample and reference contours. A squared difference is calculated specifying the dissimilarity between contours for each measurement and weighted inversely by a variance which is calculated from the set of training contours used to construct the reference. The effect of using the variance is to weight more heavily those segments in which a particular measurement is consistent over the set of training contours.[4] The various segment-by-segment measurements characterize the shape of the contours. In addition, the system computes distances based on the overall cross correlation of sample and reference contours (after time alignment) and distances based on the amount of warping requried to register the sample contours to the reference contours.

These distances are combined into an overall distance in two different ways. The first, the "overall distance" procedure, is a simple (unweighted) average over the entire set of individual distances. In the second procedure, the "selected distance" is a simple average calculated over a prespecified speaker-dependent subset of the entire set of distances. The subset is obtained as part of the training procedure by selecting those distances which are most effective in separating populations of customer and impostor utterances.

For either procedure, the combined distance is compared with a speaker-dependent threshold to determine whether to accept or reject the identity claim. The threshold distance, obtained from the training set and included in the reference data, is estimated from the overall distance distribution of distances from customer and impostor training utterances. Normally, a threshold is chosen to equalize the false (impostor) acceptance and false (customer) rejection rates. In the absence of direct knowledge of the costs of rejecting a customer or accepting an impostor, setting the threshold to give equal error rates yields the minimum cost. This is called the equal error criterion in this paper. In many real-world applications, the costs of these two types of errors would not be equal—e.g., in a banking situation the cost of rejecting a customer would be lower than the cost of accepting an impostor. In such cases, the thresbold would be adjusted appropriately.

III. EXPERIMENTAL EVALUATION

To evaluate the automatic verification system of tbe previous section, a speech data base of utterances was created. The experimental setup for creating this speech data base is shown in Fig. 2. The speech was recorded in a sound booth over conventional dialed-up telephone lines. The signal was bandlimited from 100 to 3200 Hz (the nominal telephone bandwidth) and digitized at a 10-kHz rate. Both the reference and the test utterances were processed by one of the following three transmission systems:

(*i*) Clear channel—i.e., no additional processing.
(*ii*) Adaptive differential pulse code modulation (ADPCM) coding.
(*iii*) Linear predictive vocoding (LPC).

The ADPCM coder used in this experiment was a simulation of the coder built by Bates,[7] based on the work of Cummiskey et al.[8] Figure 3 is a block diagram of the ADPCM system. Since the required sampling rate for the ADPCM coder was 6 kHz, a sampling rate conversion system was used to convert from 10 to 6 kHz at the input to the coder.[9] The signal bandwidth was reduced to 2600 Hz for the ADPCM coder by using a 100- to 2600-Hz bandpass filter in the sampling rate conversion system. In the coder, a 4-bit adaptive quantizer was used to code the difference signal ($\delta(n)$ in Fig. 3), giving an overall bit rate of 24 kbits/s for the coder. The step-size multiplier of the quantizer ranged over a 41-dB range (i.e., the ratio between the smallest step size was 114 to 1). A first-order predictor was used with a multiplier coefficient of $\alpha = 0.9375$. Signal levels were chosen so that the coder was operating at approximately the optimum range.[8]

A block diagram of the LPC vocoder is given in Fig. 4. The implementation was based on the autocorrelation method of linear prediction.[10-12] Pitch detection and voiced-unvoiced decision were performed

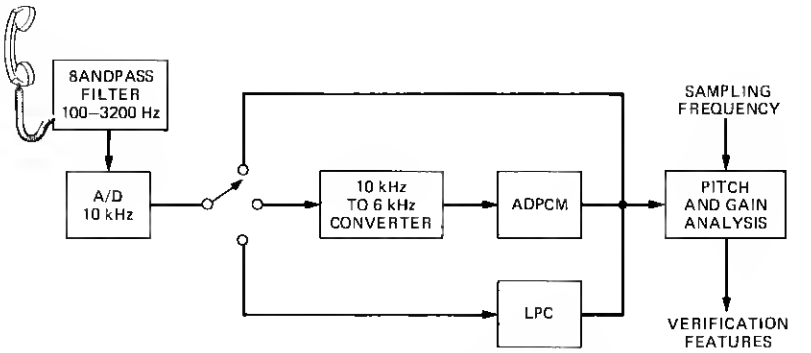

Fig. 2—Block diagram of the data collection system.

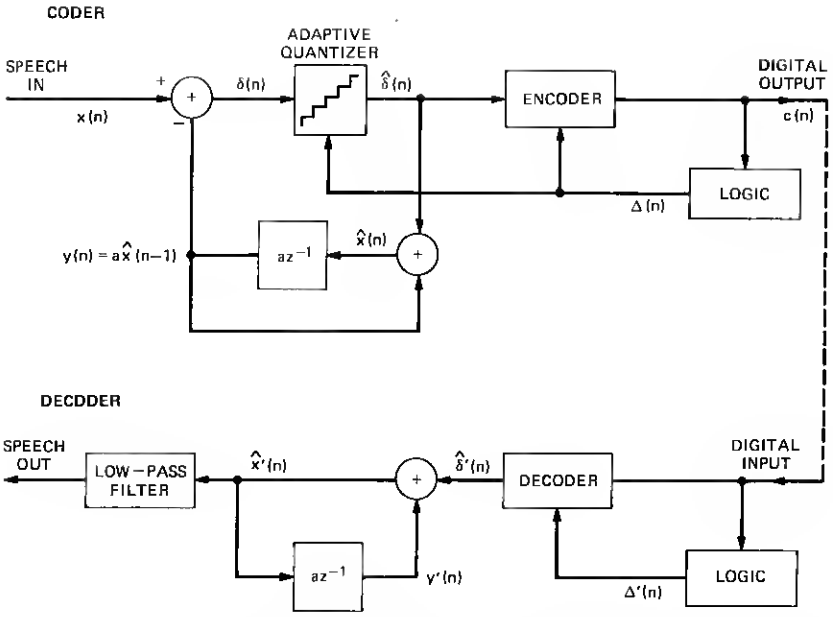

Fig. 3—Block diagram of the ADPCM coder.

using the modified autocorrelation pitch detector of Dubnowski et al.[13] A 12-pole LPC analysis was performed using a pitch-adaptive, variable frame size, at a rate of 100 frames per second.[14] No quantization of the LPC parameters was used in this experiment.

To evaluate the effects of the three transmission systems on verification accuracy, a data base was designed which included:

(*i*) Fifty recordings made by each of 20 experienced talkers (10 male and 10 female) over a period of two months. The first 10 recordings were made once a day; the remaining 40 were made twice a day (morning and afternoon). These talkers were designated "customers."

(*ii*) One recording made by each of 80 naive talkers (40 male and 40 female). These talkers were designated "impostors." There was no attempt to mimic the "customers."

Two all-voiced sentences were used in the recordings. The males used the sentence, "We were away a year ago," and the females used the sentence, "I know when my lawyer is due." In previous studies, only the first sentence was used.[2-5]

Since the automatic speaker verification system used pitch and intensity contours as features, these contours were measured once and stored on disk for later retrieval in the experiment.

3.1 *Reference construction*

For each customer and each transmission system, pitch and intensity contours from 10 of the 50 utterances were used to construct "refer-

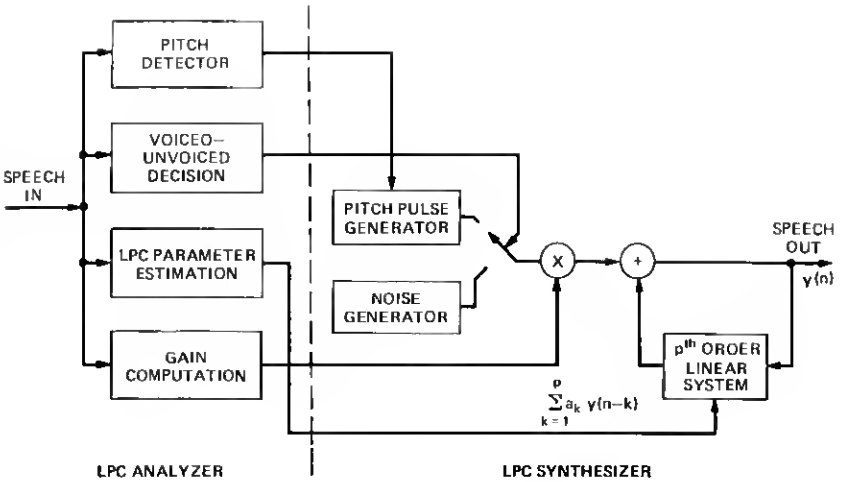

Fig. 4—Block diagram of the LPC vocoder

ence" contours. As discussed previously, in order to complete the reference construction (i.e., to get weights, thresholds, and to choose selected distances), 10 additional utterances of the 50 were used, along with the pitch and intensity contours of 15 impostors of the same sex as the customer. Two different sets of reference files were created for each customer—one obtained from the first 20 consecutive recordings of the customer (method 1) and the other obtained by using two of every five utterances recorded (method 2).

Figure 5 is a series of plots of relative cumulative frequency distributions of customer and impostor samples as a function of combined distance. Each column contains the results for two female and two male customers using the (method 1) training data. In each plot, the customer sample distributions (on the left) show the fraction of samples with distances greater than the abscissa value while the impostor sample distributions (on the right) show the fraction of samples with distances less than the abscissa value. The first column shows the results for clear channel utterances; the second column shows results for LPC vocoded utterances, and the third column shows results for ADPCM coded utterances. The decision threshold is chosen as the distance where the cumulative distributions cross (i.e., the equal error threshold), or, in the case when the distributions are separated, the point midway between the ends of the separate distributions. Since there was only a small number of training utterances, the distributions for the two types of errors are poorly defined and only a rough estimate of the decision threshold is obtained. It should be clear that the equal error threshold will vary for each pair of transmission systems being compared, as well as for different talkers. For the four talkers shown in Fig. 5, the *worst* equal error threshold indicates a 10-percent error

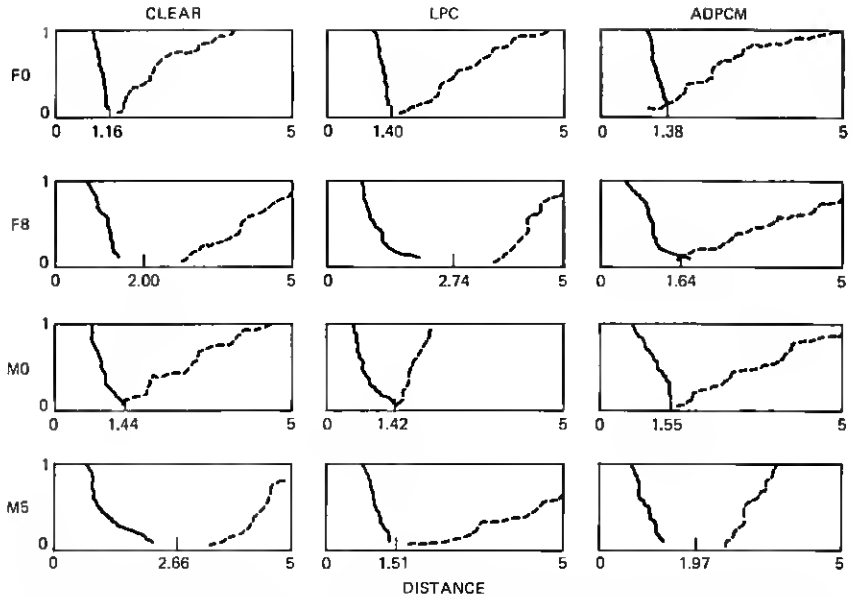

Fig. 5—Plots of the cumulative distributions of the errors for four talkers and three transmission systems based on the training set of utterances.

rate; however, for the 20 talkers, the average equal error rate over all conditions was about 1 percent. Figure 5 also shows that the value of the threshold distance varies considerably (2.5 to 1) across talkers, as does the shape of the cumulative distributions. These results tend to indicate that the training data are inadequate for obtaining a good estimate of the equal error rate threshold.

IV. RESULTS ON AUTOMATIC SPEAKER VERIFICATION

To test the automatic speaker verification system for each transmission system, each customer reference was compared to the 30 customer utterances and the 25 impostor utterances which were not used in the training set. For each set of comparisons, a Type 1 (customer rejection) and a Type 2 (impostor acceptance) error score was measured. If we denote the Type 1 error scores as E_1 and the Type 2 error scores as E_2, then E_1 and E_2 are functions of:

(*i*) The transmission system used in the training, i, where $i = 1$ denotes the clear channel, $i = 2$ denotes the LPC vocoder, and $i = 3$ denotes the ADPCM coder. The mnemonics C, V, and A are used in the plots to denote clear channel, LPC vocoder, and ADPCM coder, respectively.

(*ii*) The transmission system used in the testing, j, where $j = 1$, 2, and 3 are identical to $i = 1$, 2, and 3.

(*iii*) The training method, k, where $k = 1$ denotes training method 1 and $k = 2$ denotes training method 2.

(*iv*) The type of measurements used in the verification distance, l, where $l = 1$ is selected measurements (speaker specific), and $l = 2$ is overall measurements (speaker independent).

(*v*) The customer, m, where $m = 1$ to 10 for the 10 male customers and $m = 11$ to 20 for the female customers.

Since different sentences were used for male and female customers, results are presented separately for each subset of the customers.

To illustrate some of the results, Fig. 6 shows plots of E_1 and E_2 as a function of the training system, testing system pair (i, j), and talker (m), for selected distance measurements ($l = 1$), and training method 2 ($k = 2$). Figures 6a and 6b are E_1 scores for male customers ($m = 1$ to 10), and female customers ($m = 11$ to 20), and Figs. 6c and 6d are E_2 scores for male and female customers. A bar graph denotes the error score for each condition. The reader should note that within each group there are 10 bars, some of which are 0 indicating zero error. From this figure the following observations can be made:

(*i*) E_2 scores are significantly smaller than E_1 scores, indicating that the distance threshold obtained from the training set for equal errors ($E_1 = E_2$) was not a stable point.

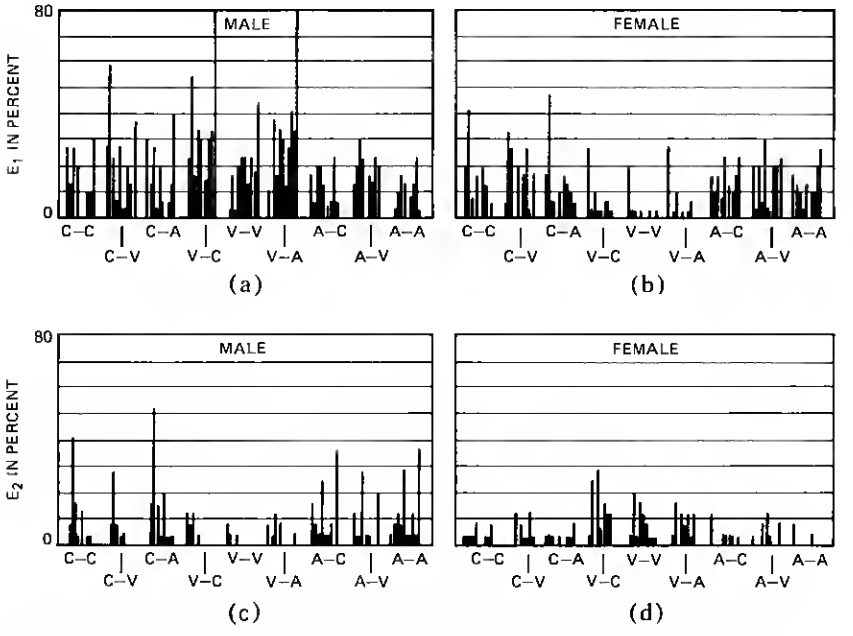

Fig. 6—Plots of E_1 and E_2 versus system pair and talker for male and female talkers (training method 2 using selected distance measurements).

(*ii*) A high degree of variability in scores (for both E_1 and E_2 and for male and female customers) exists among customers for each pair of transmission systems. This type of result was also obtained earlier by Rosenberg.[2]

(*iii*) Error scores for female customers are somewhat smaller than error scores for male customers.

(*iv*) The variability between scores for pairs of transmission systems is smaller than the variability of scores within a pair of transmission systems.

Based on the above observations, it is clear that the E_1 and E_2 scores cannot simply be averaged over customers because of the high variability among customers. Thus two data processing procedures were tried, one to use the median over customers, the other to statistically eliminate the extremes of the distribution of error scores (using a recently described method[15]) and then to average the scores of the remaining customers. Both data processing methods yielded essentially the same results and hence only the results of taking medians are presented here.

Table I gives values of the medians of E_1 and E_2 over (male and female) customers for each (i, j) pair, and for $k = 1$ and 2, and $l = 1$ and 2. Also included in this table is the median of the quantity

$$E_3 = (E_1 + E_2)/2,$$

which is the average error rate of the system. Since E_1 and E_2 were significantly different, E_3 provides a better measure of the overall performance of the verification system than either E_1 or E_2. It can be seen from Table I and statistically verified at the 0.001 level that training method 2 provides significantly better scores than training method 1, and that using selected measurements for the distance score provides significantly better scores than the overall measurements. As such, we restrict our discussion to this case only, i.e., selected measurements from training method 2.

Figure 7 is a plot of the E_3 median scores for each pair of transmissions systems. Although there is some variability in score among these systems, the variability is statistically insignificant (at the 0.01 level). Thus the major result of the testing is that the verification accuracy is relatively insensitive to the transmission system used for training and testing. This result is very different from the one obtained when verification is performed by human listeners as discussed previously. To ensure that the human verification accuracy for this new data base remained the same, the perceptual verification experiment was repeated, and the results are given in the next section.

Before the results of the perceptual experiment are described, Fig. 8 illustrates the variability of the distance threshold in the testing.

Table 1—Median over customers of the error rates

	E_1		E_2		E_3		E_1		E_2		E_3	
	male	female	male	female	male	female	male	female	male	female	male	female
C-C	16.66	11.66	20.0	8.0	17.83	14.01	23.33	16.67	8.0	4.0	17.83	9.16
C-V	25.00	10.00	16.0	8.0	24.50	14.00	28.33	20.00	2.0	2.0	16.67	13.17
C-A	15.00	11.66	20.0	8.0	18.83	15.84	21.66	15.00	6.0	2.0	14.34	7.84
V-C	35.00	18.33	8.0	16.0	23.50	20.16	35.00	15.00	4.0	4.0	19.66	12.33
V-V	25.00	5.00	8.0	8.0	15.83	7.67	28.33	6.67	4.0	4.0	16.17	5.17
V-A	30.00	11.66	6.0	8.0	19.00	12.17	36.67	13.33	4.0	4.0	20.33	9.33
A-C	18.33	18.33	16.0	6.0	23.00	13.01	23.33	13.33	4.0	4.0	19.66	9.99
A-A	20.00	10.00	8.0	4.0	22.83	12.16	26.66	13.33	4.0	2.0	17.33	11.16
A-V	18.33	16.67	12.0	6.0	22.66	13.51	30.00	15.00	4.0	4.0	23.83	10.17

(a) Overall Measurement, Training Method 1

(b) Selected Measurement, Training Method 1

	E_1		E_2		E_3		E_1		E_2		E_3	
	male	female	male	female	male	female	male	female	male	female	male	female
C-C	11.66	6.67	22.0	12.0	21.50	9.83	11.66	11.66	4.0	4.0	10.17	8.00
C-V	18.33	10.00	14.0	6.0	16.16	9.00	21.66	16.67	2.0	4.0	14.51	10.34
C-A	20.00	8.30	18.0	12.0	17.33	8.16	13.33	10.00	4.0	4.0	12.50	7.17
V-C	30.00	11.67	8.0	14.0	24.00	11.01	30.00	3.33	0.0	10.0	16.66	8.50
V-V	18.33	5.00	4.0	12.0	13.00	9.16	20.00	3.33	0.0	4.0	11.01	3.84
V-A	33.33	8.33	8.0	10.0	22.83	11.17	31.66	3.33	2.0	6.0	16.66	5.67
A-C	20.00	15.00	8.0	10.0	17.83	13.34	10.00	15.00	6.0	4.0	11.51	8.33
A-V	21.66	18.33	6.0	12.0	17.00	10.00	18.33	20.00	4.0	2.0	11.84	10.00
A-A	16.67	15.00	8.0	8.0	16.16	11.50	10.00	11.66	8.0	0.0	11.67	6.67

(c) Overall Measurement, Training Method 2

(d) Selected Measurement, Training Method 2

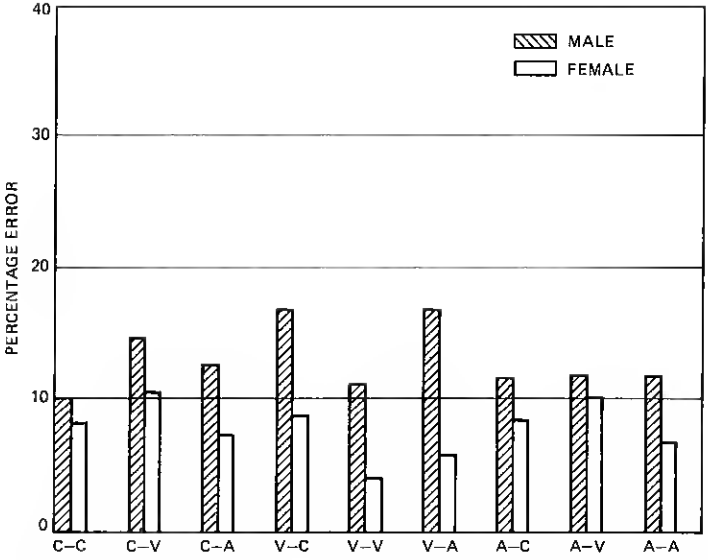

Fig. 7—Overall median error rate as a function of system pair for male and female talkers.

This figure shows the sum of the cumulative error distributions for the testing results for several different cases. Each part of the figure contains three vertical lines. The solid vertical line is the *a priori* distance threshold which gives an equal error based on the training data. The dashed vertical line is the *a posteriori* distance threshold that gives equal error based on the testing data. The dotted vertical line is the threshold that minimizes the total error E_3 for the testing data. In the ideal case, all three thresholds would be equal. However, because of the inadequacy of the training data and the unusual shapes of the cumulative error distribution, thresholds were different, as seen in this figure. The variability in both distance thresholds and error scores is fairly large (typically, between 10 and 30 percent in most cases).

V. HUMAN VERIFICATION TEST RESULTS

To verify that the human verification accuracy was strongly affected by the speech transmission system when using the telephone recordings, the experiment performed in Ref. 1 was repeated exactly. The results of these tests are given in Figs. 9 and 10. Figure 9 shows false alarm (customer rejection or Type 1 error) and miss (impostor acceptance or Type 2 error) rates for the male and female customers as a function of the pair of transmission systems used in the comparison. Only the first eight customers (female and male) were used to corre-

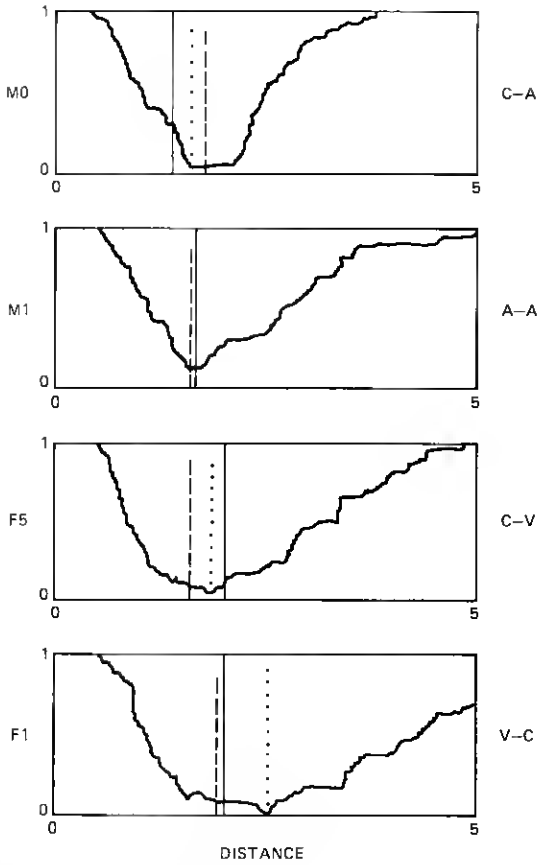

Fig. 8—Total error ($E_1 + E_2$) as a function of distance threshold for four conditions for the testing data.

spond to the eight customers used in the earlier experiment. This figure again shows the high degree of variability of the scores among customers. Thus, to combine customer scores, a median was used instead of averaging. Figure 10 shows the median false alarm rate, miss rate, and overall error rate for each pair of speech transmission systems.

The results shown in Figs. 9 and 10 are essentially identical to those reported previously,[1] namely, that the false alarm rates for transmission pairs which were the same were statistically significantly lower than for mixed systems, whereas the miss rates for homogeneous systems were larger than for mixed systems. Statistical comparisons showed that, in this experiment, the results were not statistically significantly different from those of the earlier experiment in any category.

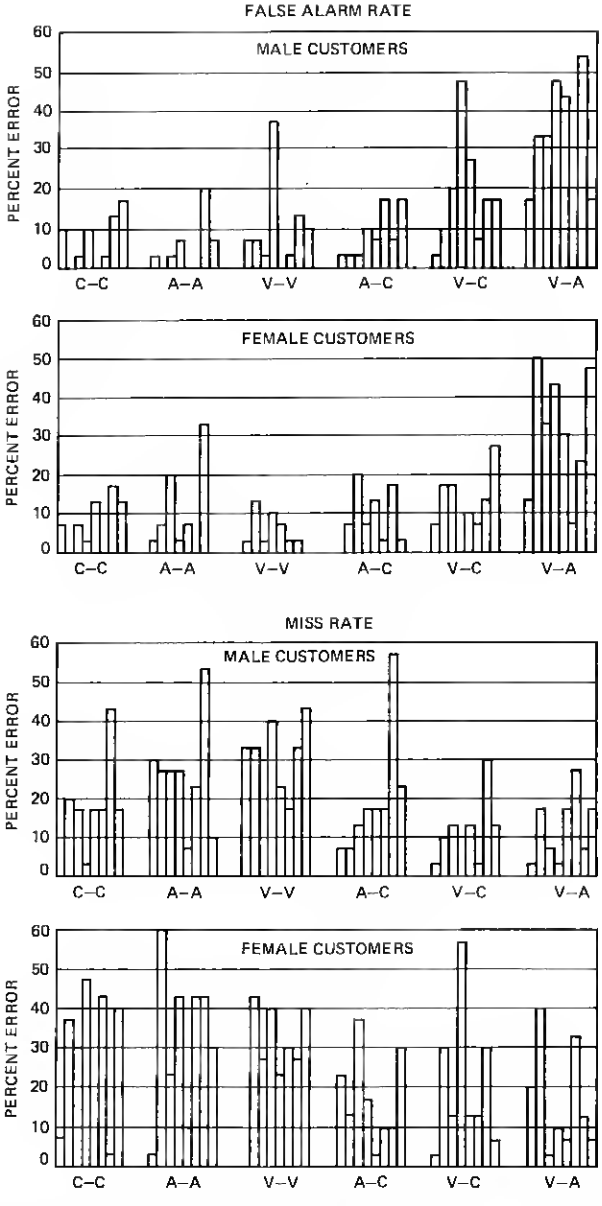

Fig. 9—False alarm and miss rates as a function of the system pair for male and female customers for the human verification experiment.

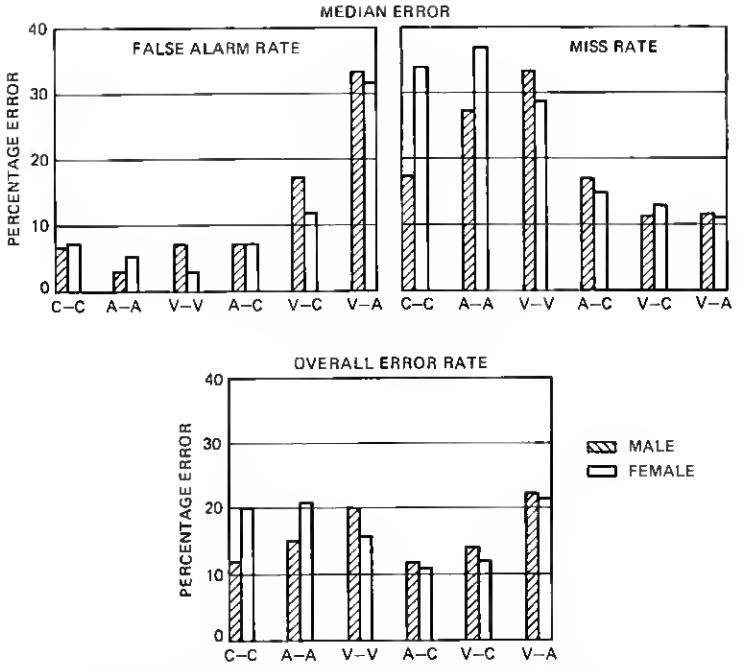

Fig. 10—Median false alarm, miss, and overall error rates as a function of the system pair for male and female customers for the human verification experiment.

VI. DISCUSSION

The major result of this investigation was the finding that, for an automatic speaker verification, accuracy was statistically insensitive to either ADPCM coding or LPC vocoding of the speech utterance for either (or both) the reference and test utterances. For this same data base, the verification accuracy by human listeners was significantly affected by the different transmission systems in the manner described in Ref. 1. The conclusion drawn from these results is that pitch and gain are reasonably robust to the distortions of ADPCM coding and LPC vocoding, thereby enabling the automatic system to be insensitive to these transmission systems.

The overall verification accuracy in this system was about 12 percent for male talkers and 8 percent for female talkers. These verification rates are comparable to those obtained by Rosenberg in a large experiment over dialed-up telephone lines.[2] As in the earlier work, considerable variability in verification scores among talkers was found, again indicating that the variability of pitch and gain for some talkers is large, and thus for these talkers other feature sets should be considered for verification. Recent unpublished investigations by Furui indicate significantly smaller (on the order of 0.5 percent) error rates

when the features are cepstral coefficient contours rather than pitch and gain. Whether such feature sets are robust to coding and vocoding remains to be investigated.

It was found in this investigation that the training methods used were inadequate for giving stable reference contours and reliable distance thresholds. This effect was previously noted by Rosenberg,[2] Furui,[16] and Furui et al.,[17] who showed that long-time variability in feature contours had to be taken into consideration to obtain stable reference data for verification.

Two other effects were noted during the course of this study. First, it was found that selected distances provided significantly better scores than overall distances. Rosenberg also noted that, once a reasonable amount of training data was obtained, selected distances were better than overall distances.[2] Thus the results here indicate that 10 training utterances are sufficient for selected distance scores to be superior to overall distance scores. The second point concerned the lower error rates for female talkers than for male talkers. Rosenberg found no statistically significant differences between verification scores for males and females.[2] Thus, this result may be due to the difference in sentence used in the verification task. If this is true, then the implication is that the test utterance chosen may provide small but consistent improvements in the verification scores.

VII. SUMMARY

In this paper, we have shown that, whereas the false alarm and miss rates for verification by human listeners are strongly affected by the pair of transmission systems used for the reference and test utterances, the false alarm and miss rates for an automatic verification system based on pitch and gain are relatively insensitive to the transmission system in the case of ADPCM coding and LPC vocoding. Although the average overall error rate for this system was around 10 percent, the robustness of pitch and gain to transmission systems makes them attractive features for automatic speaker verification systems.

REFERENCES

1. C. A. McGonegal, L. R. Rabiner, and B. J. McDermott, "Speaker Verification by Human Listeners Over Several Speech Transmission Systems," B.S.T.J., *57*, No. 8 (October 1978), pp. 2887–2900.
2. A. E. Rosenberg, "Evaluation of an Automatic Speaker Verification System over Telephone Lines," B.S.T.J., *55*, No. 6 (July–August 1976), pp. 723–744.
3. G. R. Doddington, "A Computer Method of Speaker Verification," Ph.D. Dissertation, Department of Electrical Engineering, University of Wisconsin, 1970.
4. R. C. Lummis, "Speaker Verification by Computer Using Speech Intensity for Temporal Registration," IEEE Trans. Audio and Electroacoust., *AU-21* (April 1973), pp. 80–89.
5. A. E. Rosenberg and M. R. Sambur, "New Techniques for Automatic Speaker Verification," IEEE Trans. Acoustics, Speech and Signal Processing, *ASSP-23* (April 1975), pp. 169–176.

6. A. E. Rosenberg, "Automatic Speaker Verification: A Review," Proceedings of the IEEE, *64,* No. 4 (April 1976), pp. 475–487.
7. S. L. Bates, "A Hardware Realization of a PCM-ADPCM Code Converter," M.I.T. M.S. Thesis, Dept. of Elec. Eng. and Comp. Sc., January 1976.
8. P. Cummiskey, N. S. Jayant and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J., *52,* No. 7 (September 1973), pp. 1105–1118.
9. R. E. Crochiere and L. R. Rabiner, "Optimum FIR Digital Filter Implementations for Decimation, Interpolation, and Narrow-Band Filtering," IEEE Trans. on Acoustics, Speech and Signal Proc., *ASSP-23,* No. 5 (October 1975), pp. 444–456.
10. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech,* New York: Springer-Verlag, 1976.
11. J. D. Markel and A. H. Gray, Jr., "A Linear Prediction Vocoder Simulation Based Upon an Autocorrelation Method," IEEE Trans. on Acoustics, Speech and Signal Proc., *ASSP-22,* No. 2 (April 1974), pp. 124–134.
12. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am., *50* (1971), pp. 637–655.
13. J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-Time Digital Hardware Pitch Detector," IEEE Trans. Acoust., Speech, Signal Proc., *ASSP-24* (February 1976), pp. 2–8.
14. L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," IEEE Trans. Acoust. Speech, and Signal Proc., *ASSP-25,* No. 1 (February 1977), pp. 24–33.
15. H. J. Chen, unpublished work.
16. S. Furui, "Effects of Long-Term Spectral Variability on Speaker Recognition," Paper NNN28, Acoustical Society Meeting, Honolulu, 1978.
17. S. Furui, F. Itakura, and S. Saito, "Personal Informantion in the Long-Time Averaged Speech Spectrum," Review of Elec. Comm. Lab, *23,* No. 9–10 (September–October 1975), pp. 1133–1141.